

ABSTRACT

The modernization of Indonesia's National Health Insurance Program has resulted in the changes on healthcare billing process, requiring it to adapt to more data and records. This phenomenon makes early detection of irregularities and unusual claim patterns essential for maintaining financial integrity and system efficiency. However, healthcare billing data is inherently imbalanced, high-dimensional, and largely unlabeled, rendering traditional rule-based or manual review tedious and hard to do. This study proposes a hybrid anomaly detection combining Auto-Encoder and Isolation Forest to identify atypical billing behaviors within BPJS healthcare claims, in hope to help auditors in their decision-making by surfacing records that might warrant closer examination.

The Auto-Encoder learns the latent structure of normal billing patterns and detects anomalies through reconstruction error, while the Isolation Forest isolates unusual data points based on efficient tree-based partitioning. By integrating both methods, the hybrid approach leverages the strengths of deep feature learning and distribution-based isolation to improve anomaly detection robustness.

Experiments were conducted using publicly available BPJS billing data, with preprocessing and feature engineering applied to address sparsity and high dimensionality. Due to the absence of expert-verified labels, evaluation relies on unsupervised metrics and score analysis, including reconstruction error distributions, Isolation Forest anomaly scores, threshold-based flagging (p99), and agreement rates between the models.

The result indicate that the hybrid AE-IF model consistently highlights claim records that deviate from typical billing patterns, and may warrant further expert review. While it does not classify definitive fraud, the model serves as a practical decision-support tool to prioritize suspicious or anomaly claims, improve auditing efficiency, and support more data-driven model within the national healthcare system.